



Aalborg Universitet

**AALBORG UNIVERSITY**  
DENMARK

## Awareness and Its Use in Plug and Play Process Control

Knudsen, Torben

*Published in:*  
European Control Conference 2009 - ECC'09

*Publication date:*  
2009

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Knudsen, T. (2009). Awareness and Its Use in Plug and Play Process Control. In *European Control Conference 2009 - ECC'09* (pp. 4078-4084). IEEE Press.

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

# Awareness and its use in Plug and Play Process Control

Torben Knudsen

**Abstract**—Assume a existing control system needs a new sensor or actuator due to deterioration or unsatisfactory performance for other reasons. If such a new device can be implemented in a completely automated fashion it will be a very powerful tool. This is exactly the vision for the research project “Plug and Play Process Control” P<sup>3</sup>C . The first step is to get the device on the network. The next is to decide on if it is useful and if yes for what. Methods to answer these questions are discussed and developed in this paper. The results are adaptive methods to estimate a correlation based “awareness”. These methods are successfully applied to a simulation experiment where control performance is improved. An estimate of the stochastic part of a system given the deterministic part is a necessary part of the above methods. In this paper, a prediction error method (PEM) with analytical gradient is developed for this problem.

**Index Terms**—System identification; Incremental modelling, Model validation; Plug and play process control.

## I. INTRODUCTION

The concept of “Plug and Play Process Control” (P<sup>3</sup>C) is described in [1] which can be consulted for details. A more brief descriptions follow here. Imaging a existing climate control system in a live stock stable. The farmer observes increasing signs on discomfort for the animals in a corner of the stable. The company who delivered the control system offers to install a additional temperature transducer in the corner and re-design and tune the controller to keep the set point also for the additional temperature sensor. The common way to do this would be manual and consequently expensive.

Thus the vision of the project P<sup>3</sup>C is: When a new device e.g. a sensor or actuator is plugged into a functioning control system it will identify itself and the control system will automatically become aware of the new signal, determine its usefulness and exploit it in an optimal way over time.

The objectives in this paper are to develop methods to a) “become aware of the new signal” i.e. determining if there is any relation between the new signal and the existing system and b) “determine its usefulness” i.e. measuring to what extend the new signal can be used for control.

If possible sensors and actuators are in practise chosen very fast compared to the system dynamics. For this reason and to start simple only static devices are considered in this work.

The problems treated here is new and is not covered in the literature [1]. The closest research is probably on plug in and out of *known* subsystems in a network [2].

The awareness measures developed in this paper are model based. They depend on both the present model which is

assumed known and the new part for the additional device which is unknown. The necessary batch/offline parameter estimation methods for most of the new part are developed in earlier work [1]. What is missing is the stochastic part for a new output. Recursive estimation of this part is developed in this paper as it is crucial to the awareness measures.

It seems instructive to present the notion of additional devices before discussing the awareness measures. Therefore section II presents the present and additional model parameterization. In particular the parameter estimation methods for the stochastic part is developed in section III. Next in section IV the awareness measures are discussed and developed. This is followed by a simulation example in section V demonstrating the usefulness. Finally a conclusion is given in section VI.

## II. PRESENT MODEL AND ADDITIONAL DEVICES

As this is intended for multiple input multiple output (MIMO) systems a state space (SS) model seems the best choice. In this parameterization the additional parameters will appear as new rows or columns augmented to the present matrix parameters.

### A. Present Model

The present system is assumed given in a innovation model (IM) parameterization (1).

$$x_p(t+1) = A_p x_p(t) + B_p u_p(t) + K_p e_p(t), \quad (1a)$$

$$y_p(t) = C_p x_p(t) + D_p u_p(t) + e_p(t), \quad (1b)$$

$$\text{Cov}(e_p) = R_p, \quad E(e_p(t)e_p(s)^T) = 0, \quad t \neq s \quad (1c)$$

$$u_p \in \mathcal{R}^m, \quad x_p \in \mathcal{R}^n, \quad y_p, e_p \in \mathcal{R}^l \quad (1d)$$

### B. Additional Input

In the standard SS model, input is not assumed to be noise corrupted. This means that the new input part just has to be added to the otherwise unchanged IM (2). Notice especially that  $K_p$  is unchanged. The SS IM can be divided as (2a)–(2b) where subscript  $p$  and  $a$  means present and additional respectively e.g.  $u_p$  are the inputs in the present/initial system and  $u_a$  is the additional input. Then it is only necessary to estimate  $B_a, D_a$ .

$$x_p(t+1) = A_p x_p(t) + (B_p \quad B_a) \begin{pmatrix} u_p(t) \\ u_a(t) \end{pmatrix} + K_p e_p(t) \quad (2a)$$

$$y_p(t) = C_p x_p(t) + (D_p \quad D_a) \begin{pmatrix} u_p(t) \\ u_a(t) \end{pmatrix} + e_p(t), \quad (2b)$$

$$R_p = \text{Cov}(e_p)$$

$$u_p \in \mathcal{R}^m, \quad u_a \in \mathcal{R}, \quad x_p \in \mathcal{R}^n, \quad y_p, e_p \in \mathcal{R}^l \quad (2c)$$

This work is supported by The Danish Research Council for Technology and Production Sciences.

Torben Knudsen is with Automation and Control, Department of Electronic Systems, Aalborg University, Denmark. tk@es.aau.dk

### C. Additional Output

In contrast to additional input additional output is corrupted with measurement noise. The necessary augmentation to the model is then given by (3). The additional parameters to be estimated are  $C_{ap}, D_{ap}, K_p, K_{pa}$  and the covariance  $R$ .

$$x(t+1) = A_p x(t) + B_p u_p(t) + \begin{pmatrix} K_p & K_{pa} \end{pmatrix} \begin{pmatrix} e_p(t) \\ e_a(t) \end{pmatrix} \quad (3a)$$

$$\begin{pmatrix} y_p(t) \\ y_a(t) \end{pmatrix} = \begin{pmatrix} C_p \\ C_{ap} \end{pmatrix} x(t) + \begin{pmatrix} D_p \\ D_{ap} \end{pmatrix} u_p(t) + \begin{pmatrix} e_p(t) \\ e_a(t) \end{pmatrix}, \quad R = \begin{pmatrix} R_p & R_{pa} \\ R_{ap} & R_a \end{pmatrix} = \text{Cov} \begin{pmatrix} e_p \\ e_a \end{pmatrix} \quad (3b)$$

$$u_p \in \mathcal{R}^m, x \in \mathcal{R}^n, y_p, e_p \in \mathcal{R}^l, y_a, e_a \in \mathcal{R} \quad (3c)$$

### III. PARAMETER ESTIMATION

The measurements available is assumed to be input/output both before and after the arrival of a additional device. The experimental conditions can be open loop (OL) or closed loop (CL) both with sufficient excitation.

#### A. RLS Estimates of the Deterministic Part

In previous work [1] it is shown that the deterministic part of the SS model for both additional input and output can be consistently estimated by a LS type algorithm (4) where  $\theta$  is the parameters to be estimated,  $\phi$  is the regressor vector and  $y$  is the output.

$$\hat{\theta} = \left( \sum_{t=1}^N \phi(t)\phi(t)^T \right)^{-1} \sum_{t=1}^N \phi(t)y(t) \quad (4)$$

This can be rewritten into the recursive version (5) where the forgetting factor  $\lambda$  is 1 as all data are weighted equal.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + H(t)^{-1} \phi(t)(y(t) - \hat{y}(t)), \quad (5a)$$

$$\hat{y}(t) = \phi(t)^T \hat{\theta}(t-1), \quad (5b)$$

$$H(t) = \lambda H(t-1) + \phi(t)\phi(t)^T \quad (5c)$$

#### B. Estimates of the Stochastic Part Using the RPEM Method

For a additional input the stochastic part is unchanged. For a additional output convex LS methods could only be developed for the deterministic part. Therefore PEM will be used for the stochastic part.

The starting point is then the deterministic model. All the parameters  $A, B, C$  and  $D$  in (6) is then known and only  $K$  and  $R_e$  is unknown.

$$\hat{x}(t+1) = A\hat{x}(t) + Bu(t) + Ke(t), \quad (6a)$$

$$y(t) = C\hat{x}(t) + Du(t) + e(t), \quad (6b)$$

$$\text{Cov}(e) = R_e \quad (6c)$$

The setup chosen for PEM makes it almost equivalent to the ML methods that generally is superior. The non recursive PEM method can be formulated as (7). If  $l$  is the negative likelihood and  $\theta$  includes all unknown parameters this would be the ML method. Here  $\theta$  only includes  $K$  but the minimization is done twice first with  $\Lambda = I$  and second

with  $\Lambda = \hat{R}_e$  from first minimization. If then  $\Lambda = R_e \sim \hat{R}_e$  and  $\epsilon$  is Gaussian then this would be the ML method for  $K$ .

$$\hat{\theta} = \arg \min_{\theta} l(\theta), \quad (7a)$$

$$l(\theta) = \sum_{t=1}^N \epsilon(t, \theta)^T \Lambda^{-1} \epsilon(t, \theta), \quad (7b)$$

$$\epsilon(t, \theta) = y(t) - \hat{y}(t, \theta), \quad (7c)$$

$$\hat{R}_e = \frac{1}{N} \sum_{t=1}^N \epsilon(t, \hat{\theta}) \epsilon(t, \hat{\theta})^T \quad (7d)$$

For given  $K$  the optimal one step predictor  $\hat{y}(t, \theta)$  is calculated with the KF (8) corresponding to (6).

$$\hat{x}(t+1, \theta) = A\hat{x}(t, \theta) + Bu(t) + K(y(t) - \hat{y}(t, \theta)), \quad (8a)$$

$$\hat{y}(t, \theta) = C\hat{x}(t, \theta) + Du(t) \quad (8b)$$

All iterative minimization algorithms uses the first derivative of  $l$  with respect to  $\theta$  and the best algorithms also uses the second derivative. The simplest and most numerical robust algorithms is obtained if these derivatives are calculated analytically. As seen below this is possible with  $\theta = \text{vec}(K)$ . Notice that the convention (9) is used for derivatives of vector functions with respect to vector variables.

$$f: \mathcal{R}^{n \times 1} \rightarrow \mathcal{R}^{m \times 1}, \quad \frac{\partial f(x)^T}{\partial x} \triangleq \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_m} \end{pmatrix} \in \mathcal{R}^{n \times m}, \quad (9a)$$

$$\frac{\partial f(x)}{\partial x^T} \triangleq \left( \frac{\partial f(x)^T}{\partial x} \right)^T \quad (9b)$$

As a first step the derivatives of  $l$  can be found as a function of  $\epsilon$  and its gradient.

$$\frac{\partial l(\theta)}{\partial \theta_i} = \sum_{t=1}^N \frac{\partial \epsilon(t, \theta)^T}{\partial \theta_i} 2\Lambda^{-1} \epsilon(t, \theta) \Rightarrow \quad (10)$$

$$\nabla l(\theta) = \frac{\partial l(\theta)}{\partial \theta} = 2 \sum_{t=1}^N \frac{\partial \epsilon(t, \theta)^T}{\partial \theta} \Lambda^{-1} \epsilon(t, \theta) \quad (11)$$

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j} = & 2 \sum_{t=1}^N \frac{\partial^2 \epsilon(t, \theta)^T}{\partial \theta_i \partial \theta_j} \Lambda^{-1} \epsilon(t, \theta) \\ & + 2 \sum_{t=1}^N \frac{\partial \epsilon(t, \theta)^T}{\partial \theta_i} \Lambda^{-1} \frac{\partial \epsilon(t, \theta)}{\partial \theta_j} \end{aligned} \quad (12)$$

The first sum in (12) including the product of the second derivative and  $\epsilon$  will tend to zero as  $\theta \rightarrow \theta_0$  and  $N \rightarrow \infty$  because then the innovation  $\epsilon(t, \theta_0)$  will be independent

of  $\hat{y}(t, \theta_0)$  and thus also of its derivatives  $\frac{\partial^2 \hat{y}(t, \theta_0)^T}{\partial \theta_i \partial \theta_j} = -\frac{\partial^2 \epsilon(t, \theta_0)^T}{\partial \theta_i \partial \theta_j}$ .

Then a good approximation for the second derivative is:

$$H(\theta) = 2 \sum_{t=1}^N \frac{\partial \epsilon(t, \theta)^T}{\partial \theta} \Lambda^{-1} \frac{\partial \epsilon(t, \theta)}{\partial \theta^T} \sim \nabla^2 l(\theta) = \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta^T}$$

Now only the gradient of  $\epsilon$  is needed. It is given by the gradient of the one step predictor (13).

$$\epsilon(t, \theta) = y(t) - \hat{y}(t, \theta) \Rightarrow \frac{\partial \epsilon(t, \theta)}{\partial \theta} = -\frac{\partial \hat{y}(t, \theta)}{\partial \theta} \quad (13)$$

Recall that  $\theta = \text{vec}(K)$  i.e. the vector of stacked columns of  $K$ . The gradient of the output prediction can then be derived from the KF (8). The result is (14a) where  $I_n$  in (14c) is the identity matrix of size  $n$ .

$$\frac{\partial \hat{x}(t+1, \theta)}{\partial \theta^T} = A \frac{\partial \hat{x}(t, \theta)}{\partial \theta^T} - K \frac{\partial \hat{y}(t, \theta)}{\partial \theta^T} + I(\epsilon(t, \theta)), \quad (14a)$$

$$\frac{\partial \hat{y}(t, \theta)}{\partial \theta^T} = C \frac{\partial \hat{x}(t, \theta)}{\partial \theta^T}, \quad (14b)$$

$$I(\epsilon(t, \theta)) \triangleq \epsilon(t, \theta)^T \otimes I_n \\ = [\epsilon(t, \theta)_1 I_n \quad \epsilon(t, \theta)_2 I_n \quad \cdots \quad \epsilon(t, \theta)_p I_n] \quad (14c)$$

The stability of (14a) is the same as for the KF i.e. it is stable if the eigen values of  $A - KC$  is within the unit circle which they are if the system is observable.

For the iterative batch minimization the Levenberg-Marquardt [3], [4] method (15) is used. Convergence of this algorithm is basically obtained by reducing  $\delta$  for each step where  $l$  decreases and increasing  $\delta$  if  $l$  does not decrease.

$$\hat{\theta}_k = \hat{\theta}_{k-1} - \left( \delta I_{np} + \nabla^2 l(\hat{\theta}_{k-1}) \right)^{-1} \nabla l(\hat{\theta}_{k-1}) \quad (15)$$

The corresponding recursive version is (16) where the notation is simplified by leaving out some of the dependency on  $\theta$ . Again the forgetting factor  $\lambda$  is 1.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + H(t)^{-1} \psi(t) \epsilon(t), \quad (16a)$$

$$\epsilon(t) = y(t) - \hat{y}(t), \quad \psi(t) = \frac{\partial \hat{y}(t, \theta)^T}{\partial \theta}, \quad (16b)$$

$$\hat{R}_e(t) = \begin{cases} \lambda \hat{R}_e(t-1) + (1-\lambda) \epsilon(t) \epsilon(t)^T, & \lambda < 1 \\ (1 - \frac{1}{t}) \lambda \hat{R}_e(t-1) + \frac{1}{t} \epsilon(t) \epsilon(t)^T, & \lambda = 1 \end{cases} \quad (16c)$$

$$H(t) = \lambda H(t-1) + \psi(t) \hat{R}_e(t)^{-1} \psi(t)^T \quad (16d)$$

There are a number of options and choices to make when combining (5) and (16) into a full recursive solution. These rather technical details are left out in this paper.

#### IV. AWARENESS MEASURES

So far awareness has not been defined precisely. The purpose is that: Awareness should measure the potential usefulness for control.

An obvious measure is the “correlation” between a new actuator and the present sensors or between a new sensor and the present actuators. If there is a high correlation

it can probably be exploited by the controller while low correlation probably makes it difficult. There are however some difficulties with correlation. How exactly should it be measured and interpreted?

##### A. Linear correlation

Correlation can be measured by traditional *model free methods* or by the *relative reduction in modelling error*. Linear correlation between two stochastic variables is defined by (17). In accordance with this definition the model free estimator is given by (18) where  $\hat{\mu}$  and  $\hat{\sigma}$  are standard estimates.

$$\rho_{xy} \triangleq \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}, \quad (17)$$

$$\mu_x \triangleq E(x), \quad \sigma_x \triangleq \sqrt{E[(x - \mu_x)^2]}$$

$$\hat{\rho}_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\hat{\sigma}_x \hat{\sigma}_y} \quad (18)$$

Linear correlation measured by the relative reduction in modelling error is based on linear models as follows. Define the affine projection of  $y$  on  $x$  as (19). For  $m \geq 3$  the correlation (20) is between a scalar  $y$  and the best linear combination of “regressors”  $(x_1 \dots x_m)^T$  which is called the multiple squared correlation often denoted  $R^2$ . For  $m = 2$  where  $x = (1 \ x_2)^T$  the linear correlation (17) is equivalent to (20). Notice that in (19)  $x_1$  is always 1 to estimate the “level”. With this definition  $\tilde{y}(\theta^*, 1)$  means the projection on 1 such that  $\tilde{y}(\theta^*, 1) = \theta^* = \mu_y$ . An estimate based on (20) is then obtained by substituting  $\tilde{y}(\theta^*, x) = \hat{\theta}_{LS}^T x$  where  $\hat{\theta}_{LS}$  is the standard linear regression estimate and by exchanging all the mean square errors in (20) with corresponding average square errors.

$$\tilde{y}(\theta, x) \triangleq \theta^T x = (\theta_1 \quad \dots \quad \theta_m) \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}, \quad \left. \begin{aligned} x_1 &\triangleq 1, \ y \in \mathcal{R}, \ x, \theta \in \mathcal{R}^m, \\ \theta^* &\triangleq \arg \min_{\theta} E[(y - \tilde{y}(\theta, x))^2] \end{aligned} \right\} \Rightarrow \quad (19)$$

$$\rho_{xy}^2 = \frac{E(y - \mu_y)^2 - E(y - \tilde{y}(\theta^*, x))^2}{E(y - \mu_y)^2} \quad (20)$$

Another useful measure is the partial correlation. Divide  $x$  and  $\theta$  into the first  $m-1$  component and then the last component as in (21). The partial correlation between  $w$  and  $y$  given  $v$  is then defined by (22).

$$v \triangleq \begin{pmatrix} x_1 \\ \vdots \\ x_{m-1} \end{pmatrix}, \quad w \triangleq x_m \Rightarrow x = \begin{pmatrix} v \\ w \end{pmatrix}, \quad (21)$$

$$\rho_{wy|v} \triangleq \frac{E[(w - E(w|v))(y - E(y|v))|v]}{\sqrt{V(w|v) V(y|v)}} \quad (22)$$

The partial correlation can also be expressed by relative reduction in model errors. This is done by similar calculations as for (20) and gives (23) where the first error is when only  $v$  is used in the regression model and the last error is when

both  $v$  and  $w$  i.e. all  $x$  is used. The estimate for the partial correlation based on (23) is similar to the estimate based on (19).

$$\rho_{wy|v}^2 = \frac{E(y - \check{y}(\theta_v^*, v))^2 - E(y - \check{y}(\theta_x^*, x))^2}{E(y - \check{y}(\theta_v^*, v))^2} \quad (23)$$

### B. Non linear correlation for signals in stochastic systems

It is well known that non linear dependence is not necessarily seen in linear correlation. If e.g.  $x \in N(0, \sigma^2)$ ,  $y = x^2$  then  $y$  is given by  $x$  but still  $\rho_{xy} = 0$ . Such non linear dependence can be included by using the variance reduction correlation measure and changing the model  $\check{y}(\theta, x)$  to be nonlinear in  $x$ . For the type of nonlinearity there are many possibilities e.g. polynomials or smoothing techniques [5] can be chosen.

If  $x$  and  $y$  is two stochastic processes then it is possible that the correlation  $\rho_{xy} = 0$  even though one signal is given by the other e.g. if  $x(t)$  is white noise and  $y(t) = x(t-1)$ . To measure this type of correlation cross correlation functions (24) is used as they cover all lags between  $x$  and  $y$ .

$$C_{xy}(t, s) = \rho_{x(t)y(s)} \quad (24)$$

As a generalization of the above development a new suggestion for a measure of non linear correlation in stochastic systems is the following: Use the variance reduction techniques and expand the model class used for  $\check{y}$  from non linear static models to non linear dynamic models. Then it is not necessary to calculate the cross correlation function because the dynamics is included in the model.

For simpler notation assume that presently there are one input  $u_p$  one output  $y_p$ . An additional input  $u_a$  or output  $y_a$  can then be added. The combination of present and additional signal is then a two dimensional signal which is called new.

$$y_n \triangleq \begin{pmatrix} y_p \\ y_a \end{pmatrix}, \quad u_n \triangleq \begin{pmatrix} u_p \\ u_a \end{pmatrix} \quad (25)$$

It is also useful to define a collection of signals as in (26) where  $X^t$  refers to signal  $x$  from a starting to a end time  $t$ . Often the end time is  $t-1$  and is then sometimes omitted.

$$X \triangleq X^{t-1} \triangleq \begin{pmatrix} x(t-1) \\ x(t-2) \\ \vdots \\ x(1) \end{pmatrix} \quad (26)$$

With this notation different model errors as e.g. (27) can be specified where  $\hat{y}_a(t|U_p, Y_p)$  is the best prediction of the additional output  $y_a(t)$  at time  $t$  given all present inputs  $u_p$  and outputs  $y_p$  from the start and up to and including time  $t-1$ .

$$\tilde{y}_a(t|U_p, Y_p) \triangleq y_a(t) - \hat{y}_a(t|U_p, Y_p) \quad (27)$$

The model or prediction error (27) has originated from the above discussion on how to define and measure awareness in a ‘‘correlation’’ like fashion. However, the error (27) also has a very relevant interpretation in terms of control.

If a reference for  $y_a(t+1)$  is known at time  $t$  the equation (28) can in principle be solved for  $u_p(t)$  which is then a function of  $u_p$  up till time  $t-1$ ,  $y_p$  up till time  $t$  and the reference  $y_{a,r}(t+1)$  i.e. a minimal variance feedback controller. The error (27) is then the smallest reference tracking error which can be obtained using the *given* variables which in this case are  $u_p$  and  $y_p$ .

$$\left. \begin{aligned} \hat{y}_a(t+1|U_p^t, Y_p^t) &= y_{a,r}(t+1) \Rightarrow \\ u(t) &= f(U_p^{t-1}, Y_p^t, y_{a,r}(t+1)) \end{aligned} \right\} \Rightarrow \quad (28)$$

$$\begin{aligned} y(t+1) - y_{a,r}(t+1) &= y(t+1) - \hat{y}_a(t+1|U_p^t, Y_p^t) \\ &= \tilde{y}_a(t+1|U_p^t, Y_p^t) \end{aligned} \quad (29)$$

Based on the above the awareness measures in (30) are suggested. Some remarks are in place:

- 1) The awareness (30a) measures how well the additional sensor can be controlled by the present actuator in closed loop (CL) provided it is controllable. If it is not controllable a positive awareness could be solely do to  $Y_a$  i.e. auto correlation. In this case there would be no way to control the additional sensor. This situation imply a zero second awareness measure (30b). A positive awareness could also be solely do to  $y_p$  such that  $y_a$  cannot be controlled independently of  $y_p$ .
- 2) The awareness (30b) measures how well the additional sensor can be controlled by the present actuator in open loop (OL).
- 3) The awareness (30c) measures how much the additional actuator can *add* to the control of the present sensor in closed loop (CL).

$$\begin{aligned} \rho_{U_p Y_p Y_a, y_a}^2 &= \\ \frac{E[y_a(t) - E(y_a(t))]^2 - E[y_a(t) - \hat{y}_a(t|U_p, Y_p, Y_a)]^2}{E[y_a(t) - E(y_a(t))]^2} \end{aligned} \quad (30a)$$

$$\begin{aligned} \rho_{U_p, y_a}^2 &= \\ \frac{E[y_a(t) - E(y_a(t))]^2 - E[y_a(t) - \hat{y}_a(t|U_p)]^2}{E[y_a(t) - E(y_a(t))]^2} \end{aligned} \quad (30b)$$

$$\begin{aligned} \rho_{U_a y_p | U_p Y_p}^2 &= \\ \frac{E[y_p(t) - \hat{y}_p(t|U_p, Y_p)]^2 - E[y_p(t) - \hat{y}_p(t|U_p, Y_p, U_a)]^2}{E[y_p(t) - \hat{y}_p(t|U_p, Y_p)]^2} \end{aligned} \quad (30c)$$

The above ‘‘theoretical’’ awareness measures are based on models not data. Like for correlations the squared values  $\rho^2$  will be between 0 and 1. To use these measures the various versions of the function  $\hat{y}(t|X)$  must be found. This is possible for linear systems e.g. by using a KF. However, all parameters in the full model must be known. This is why parameter estimation of additional parameters is crucial.

Given a linear model there is a map from parameters to the awareness measures above. This will of course be a good solution for a correct model, but the sensitivity to modelling errors will be large.

A more robust alternative chosen here is to substitute the expected squares in (30) with corresponding time average

squares based on measured input and output. The filters used to generate the various  $\hat{y}$ 's is then based on the recursively estimated parameters.

Notice that when a model including the new device is just starting to adapt the awareness measure estimate might very well be negative.

The awareness measure suggested above is of the *variance reduction using dynamical models* type. It has many advantages e.g. calculations almost covered by the adaptive parameters estimation, covers both static and dynamic systems both linear and nonlinear in inputs and gives one simple measure in contrast to cross correlation functions. The main disadvantage is the sensitivity to modelling errors.

A additional device can have low awareness of the above type because of low control potentials even though it has a statistical significant model improvement. To also measure this the p-value  $p_f$  in a F-test [6, p. 509] for all additional parameters equal zero are included.

## V. SIMULATION EXAMPLE

To show an example where there is a expected effect of a additional output measurement a system has been constructed. The deterministic part is a zero order hold sampled version of the 2 order continuous time system (31).

$$\dot{x}(t) = Ax(t) + Bu(t), \quad (31a)$$

$$y(t) = Cx(t) + Du(t), \quad (31b)$$

$$\left[ \begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] = \left[ \begin{array}{cc|cc} -0.1 & -0.01 & 0.1 & 0.1 \\ 0.01 & -0.1 & 0 & 0.1 \\ \hline 1 & 0 & 0 & 0.1 \end{array} \right] \quad (31c)$$

Using the common notation used e.g. in [7] this gives the discrete time system (32) where the stochastic part also is chosen.

$$\left[ \begin{array}{c|c} A_p & B_p \\ \hline C_p & D_p \end{array} \right] = \left[ \begin{array}{cc|cc} 0.9048 & -0.0090 & 0.0952 & 0.0947 \\ 0.0090 & 0.9048 & 0.00047 & 0.0956 \\ \hline 1 & 0 & 0 & 0.1 \end{array} \right] \quad (32a)$$

$$R_w = \begin{pmatrix} 0.0100 & 0.0050 \\ 0.0050 & 0.0100 \end{pmatrix}, R_v = 0.01, R_{wv} = \underline{0} \quad (32b)$$

$$K_p = \begin{pmatrix} 0.5381 \\ 0.2597 \end{pmatrix}, R_e = 0.0248 \quad (32c)$$

This system is constructed such that the second state is difficult to observe. The condition number for the observability matrix is 201. This improves to 2.62 when introducing the additional output (33) because the second state now is included in the output.

$$\left[ \begin{array}{c|c} C_a & D_a \end{array} \right] = \left[ \begin{array}{cc|cc} 1 & 1 & 0.1 & 0.1 \end{array} \right] \quad (33a)$$

$$R_a = 0.1, R_{a\bullet} = R_{\bullet a} = \underline{0} \quad (33b)$$

To test the system in CL and to illustrate the potentials to improve control performance a controller is included. The LQ controller design (34) has been chosen such that the potentials of the approach is exposed. This is done by assuming a quit fast controller that aims to minimize both

states where the last state initially is hard to observe. As the state is not directly measured the optimal solution is to use the state estimate as given in (35).

$$F_o = \min_F \lim_{N \rightarrow \infty} \sum_{t=1}^N x(t)^T Q_x x(t) + u(t)^T Q_u u(t),$$

$$u(t) = -F_o x(t), Q_x = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, Q_u = 10^{-4} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (34)$$

$$u(t) = -F_o \hat{x}(t) = -F_o E(x(t)|Y^{t-1}),$$

$$F_o = \begin{pmatrix} 9.2613 & -9.2564 \\ 0.1464 & 9.3077 \end{pmatrix} \quad (35)$$

The simulation shown in figure 1 starts with 500 samples with CL operation where only the first output is used for the KF and thus for the controller. Then comes 1000 samples with external excitation added to make the new output part identifiable. After this period the KF is updated with the additional output and model as it is at time 1500. The last 500 samples is then in CL with the new controller and with out external excitation.

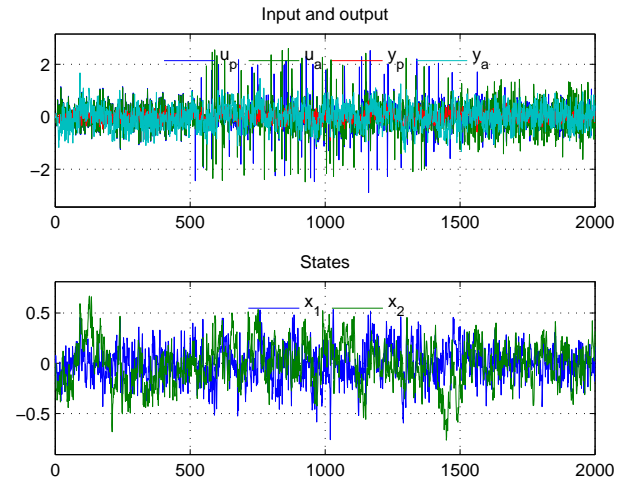


Fig. 1. Input, states and output for the simulation example. For details see text.

At time 500 the recursive calculation of additional parameters and awareness is initialized. As there are no external excitation from time 1500 the calculations can be stopped as the model will not improve from this time. After the initial phase the parameters seems to be in a process of converging. The convergence does not seem quit finished at time 1500.

The test based awareness measure shows a statistical significant relation between the new measurement and the present model after less than 100 samples according to figure 3. The correlation based awareness related to CL control of the additional output finds a value around 0.25 after 200 samples whereas the one related to improvement in CL control of the present output has only just exceeded 0 after the excitation ends. The awareness related to OL control of the new output decreases to -0.13 because the model is really inferior to a average value when it comes to pure simulation. Notice that after the external excitation

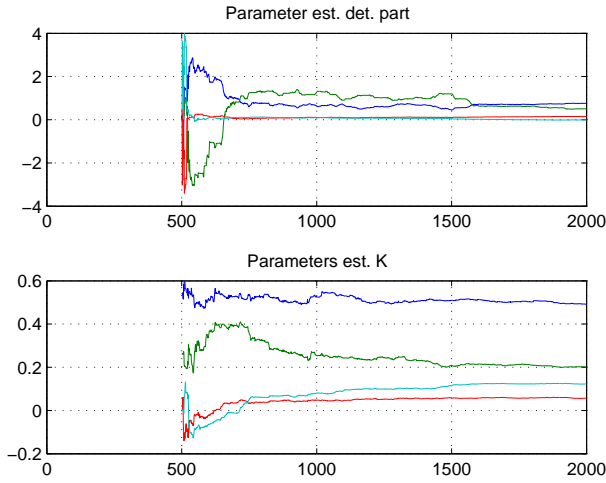


Fig. 2. Recursive parameter estimates initialized at time 500.

ends the parameter are slowly drifting and in particular the awareness related to CL control off the new output is decreasing which can be expected as the model cannot improve without excitation.

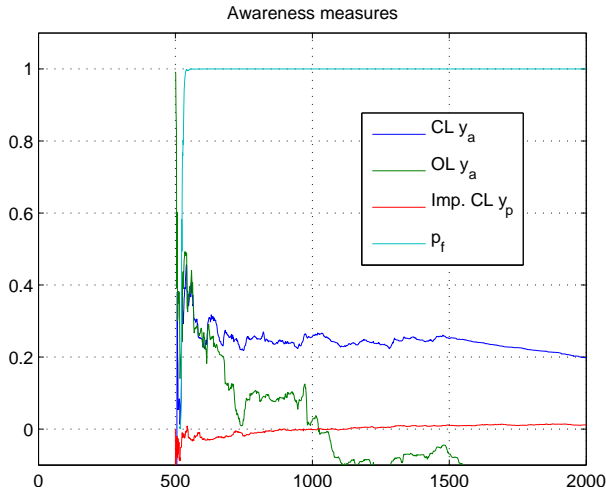


Fig. 3. Awareness measure initialized at time 500.

If the control objectives are directly related to each output a condition for switching controller could be that all CL awareness measures should be larger than zero perhaps plus some threshold. In general the control objectives are related to the states and only sometimes the objectives can be directly specified from output and input only. Still, in general it is reasonable to assume that there is a strong relation between control objectives and the outputs. Consequently, the most reasonable switch condition is CL awareness measures above zero. This is the reason for choosing time 1500 here.

To verify that the methods can improve the control performance various measures are listed in table I. The first six rows are mean squares of the signals and the next two rows  $p_u$  and  $p_x$  are the input and state part of the control

	CL pres. model	CL pres. model with excit.	CL new model	CL new/ CL pres.
$u_1$	0.162	0.352	0.158	0.971
$u_2$	0.152	0.359	0.337	2.217
$x_1$	0.023	0.040	0.018	0.779
$x_2$	0.058	0.049	0.029	0.502
$y_1$	0.026	0.046	0.023	0.870
$y_2$	0.187	0.196	0.154	0.825
$p_u$	3.142e-05	7.108e-05	4.943e-05	1.574
$p_x$	0.082	0.089	0.047	0.581
total	0.082	0.089	0.047	0.581

TABLE I  
COMPARISON OF RMS FOR SIGNALS AND CONTROL PERFORMANCE  
BETWEEN THE THREE SIMULATED INTERVALS.

performance and the last row is the total control performance. The most important observation is that the total control performance is decreased to roughly 0.6 of what it was which shows that the combination of a extra output the estimated model for it and the selected switching time improved the control. The details are that the second input doubles in MS. Both states decrease in MS especially the second which was the one that were difficult to observe. Also both outputs especially the last decrease in MS.

## VI. CONCLUSION

This paper considers the situation where a additional sensor or actuator is added to a control system with the purpose of improving the performance in a automatic way. Correlation based awareness measures are developed which tells if there is a potential gain by using the new device. Also rules for when to switch in the new device has been suggested. These are the first methods in a new area of research and further improvement should be pursued. The awareness measures are model based. Therefore it is necessary to update the present system with the part covering the new device. For the deterministic part this paper builds on previous work. For the stochastic part a prediction error method has been developed with a explicit analytical gradient. A simulation example shows that the methods can be applied with succes.

## REFERENCES

- [1] T. Knudsen, K. Trangbaek, and C. S. Kallesøe, "Plug and play process control applied to a district heating system," in *17th IFAC World Congress (IFAC'08)*, IFAC, 2008, pp. 325–330.
- [2] R. J. P. C. Kambhampati and F. J. Uppal, "Reconfiguration in plug-and-play network of embedded systems," in *Mechatronic Systems, Volume 4, Part 1*, vol. 4, IFAC. Ruprecht-Karls-University, Germany: IFAC-PapersOnLine, 2006, pp. 1065–1070, 4th IFAC Symposium on Mechatronic Systems (2006).
- [3] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Indust. Appl. Math.*, vol. 11, no. 2, pp. 431–441, June 1963.
- [4] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge University Press, 1992.
- [5] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. Chapman & Hall, 1990.
- [6] L. Ljung, *System Identification, Theory for the User*, 2nd ed., ser. Prentice-Hall information and system sciences series. Prentice-Hall, 1999.
- [7] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*. Prentice-Hall, 1996.